Mammographic Ellispe Modelling for Risk Estimation

George, Minu; Denton, Erika; Zwiggelaar, Reyer

Published in:
Procedia Computer Science

DOI:
10.1016/j.procs.2016.07.020

Publication date:
2016

Citation for published version (APA):
George, M., Denton, E., & Zwiggelaar, R. (2016). Mammographic Ellispe Modelling for Risk Estimation. *Procedia Computer Science*, *90*, 163-168. https://doi.org/10.1016/j.procs.2016.07.020

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 90 (2016) 163 – 168

Procedia Computer Science

International Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016, 6-8 July 2016, Loughborough, UK

Mammographic Ellipse Modelling for Risk Estimation

Minu George[a,*], Erika Denton[b], Reyer Zwiggelaar[a]

[a]*Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, UK*
[b]*Department of Radiology, Norfolk and Norwich University Hospital, Norwich, NR4 7UY, UK*

Abstract

It has been shown that breast density and parenchymal patterns are significant indicators in mammographic risk assessment. In addition, studies have shown that the sensitivity of computer aided tools decreases significantly with increase in breast density. As such, mammographic density estimation and classification plays an important role in CAD systems. In this paper, we present the classification of mammographic images according to breast parenchymal structures through a multi-scale ellipse blob detection technique. Our classification is based on classifying the mammographic images of the MIAS dataset into high/low risk mammograms based on features extracted from a blob detection technique which is based on breast tissue structure. In addition, it evaluates the relation between the BIRADS classes and low/high risk mammograms. Results demonstrate the probability of estimating breast density using computer vision techniques to improve classification of mammographic images as low/high risk.

Peer-review under responsibility of the Organizing Committee of MIUA 2016

Keywords: Breast density modelling; blob and ellipse detection;breast density classification ; mammographic risk estimation;

1. Introduction

Breast cancer is considered as a major health problem among women in western countries .It is estimated that one among eight women in their life span has the chance of developing breast cancer. Breast tissue density and parenchymal pattern structures are indicative of there being a high risk of developing breast cancer and there is a direct relation with the probability of developing breast cancer[1,2,3]. The recent studies have shown that the sensitivity of computer aided tools is significantly reduced with the increase in density of the breast tissue. A few examples of mammographic images are illustrated in Fig. 1 clarifying various types of tissue densities as per the Breast Imaging Reporting and Data System (BIRADS) density classification. The manual assessment of breast density and/or parenchymal patterns is covered in various classification schemes[1,2,4,5] which are all correlated[6].The automatic segmentation of mammographic tissue and the related (BIRADS) density classification of the mammograms has been an active research area. In general this has concentrated on the modelling and detection of dense tissue[7,8,11]. The review by He et al.[10] demon-

* Corresponding author.
E-mail address: mig24@aber.ac.uk

Peer-review under responsibility of the Organizing Committee of MIUA 2016
doi:10.1016/j.procs.2016.07.020

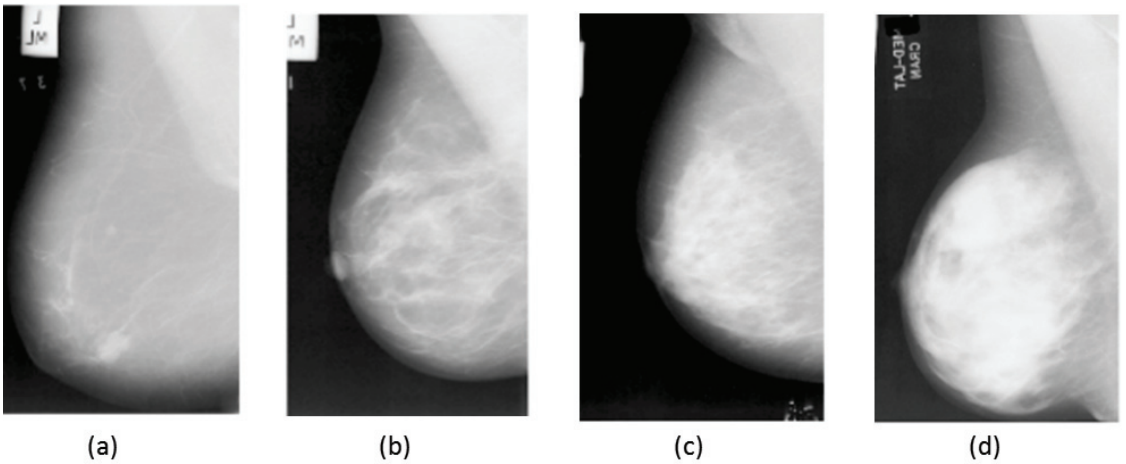

Fig. 1. Some example mammograms showing various BIRADS density classification based on Birads 4th edition. (a) BIRADS 1 (entirely fatty breast tissue), (b) BIRADS 2 (scattered fibro-glandular density), (c) BIRADS 3 (heterogeneously dense breast obscuring small masses), and (d) BIRADS 4 (extremely dense breast lowering the sensitivity of mammography).

strated all recent approaches in finding the correlation between the breast density and its classification methods in risk assessment. In [10] they showed that the main approaches in the literature use techniques like thresholding, clustering, and statistical model building.

In this paper, we demonstrate the novel approach of modelling the breast tissue : i.e. the modelling of distribution of dense tissue and fatty tissue and how this can be exploited to provide the BIRADS density classification. This work is closely related to that of Chen et al. [9], which developed a mammographic segmentation approach using topographic maps with an emphasis on the modelling and distribution of bright blobs in mammographic images. In section 2, the creation of gLoG kernel is briefly described. Later in section 3, the results obtained from the blob detection technique is discussed [18]. In section 4, the classification of the mammographic images from the Mammographic Image Analysis Society (MIAS) database into BIRADS classes based on the features obtained from the multi-scale blob detection technique is described. In addition the classification of mammographic images into low/high risk images based on blob detection is explained. Lastly section 5 discusses the results obtained and future work.

2. Blob Detection from Mammogram Images

In computer vision, the regions or points which are identified as either darker or brighter than the local area are referred as blobs. The higher the contrast between the region and the surrounding tissue and the more blob-like the region is the higher the blob probability should be. Traditional approaches have concentrated on the detection of circular objects for which approach such as the Hough transform [12] and Laplacian of Gaussian (LoG) filtering [13] have been used. However, a lot of regions in natural images are not truly circular and as such ellipse detection might be a more appropriate approach. This leads to the introduction of generalized Laplacian of Gaussian (gLoG) based approach, which allows for the detection of asymmetric regions and includes directional estimation for such regions [14]. The gLoG approach uses a set of orientation and scale dependent kernels. As with the LoG approach, the image, $f(x, y)$ is processed with a Gaussian defined by

$$\nabla^2[G(x, y) * f(x, y)] = \nabla^2[G(x, y)] * f(x, y) \tag{1}$$

where

$$\nabla^2 G(x, y) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} \tag{2}$$

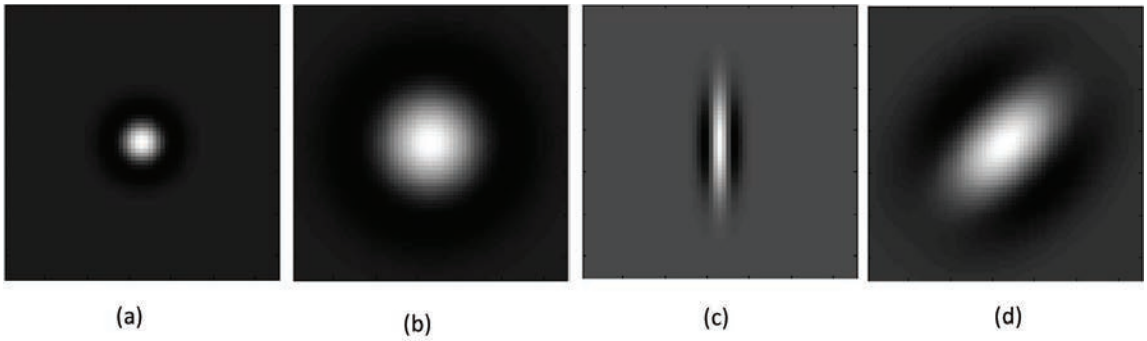

Fig. 2. A set of gLoG example kernels, where (a) $\{\sigma_x = 4, \sigma_y = 4\}$, (b) $\{\sigma_x = 10, \sigma_y = 10\}$, (c) $\{\sigma_x = 8, \sigma_y = 2, \theta = \pi/2\}$, and (d) $\{\sigma_x = 12, \sigma_y = 8, \theta = \pi/4\}$.

and

$$G(x, y) = A.e^{-(ax^2+2bxy+cy^2)} \tag{3}$$

In Equation3, the coefficients a, b, and c controls the shape and orientation of the kernel G(x, y). the parameters are determined by

$$a = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2} \tag{4}$$

$$b = -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2} \tag{5}$$

$$c = \frac{\sin^2\theta}{2\sigma_x^2} + \frac{\cos^2\theta}{2\sigma_y^2} \tag{6}$$

where σ_x, σ_y and θ are respectively the standard deviation in the horizontal direction, the standard deviation in the vertical direction, and the orientation of the Gaussian kernel. The two σ_x, σ_y values determine the scale/width of the Gaussian kernel, and as there are two values this can now be asymmetrical.Some typical kernels can be found in Fig. 2. This illustrates a few possible varieties of kernels obtained while changing the orientation and scale parameters.

3. Data and Experimental Results

For this initial study we have used the MIAS dataset[15] , which contains 322 digitised MLO mammograms. The database consists of left and right MLO mammograms of 161 women. The spatial resolution of images is $50\mu m \times 50\mu m$ quantised to 8 bits. All the mammograms were classified according to the BIRADS density classification scheme[4] by expert breast radiologists[16]. As an initial step, all the images are pre-processed and the breast area is separated from the background region and from the pectoral muscle. Blob-like structures detected outside the breast area are also removed. Subsequently, all the mammograms in the dataset were processed with Equation. 1, using a variety of $\{\sigma_x, \sigma_y, \theta\}$ kernels. A few typical examples can be found in Fig. 3. This shows that different size and orientation structures are localized and enhanced by the different kernels. Based on the results, we identify the local maxima which indicate the central location of each blob-like structure and thresholding is used to select the regions with higher probability. This approach is done to extract blob-like fatty and dense structures from the mammogram images. In order to maintain the computational cost, repeated filtering of the mammographic image with a filter of increasing size across scale space is avoided. Instead downsampling of the mammographic image is performed giving more stability to the gLoG kernel.

To exhibit the working of blob detection on various density types, mammograms from BIRADS I to BIRADS IV are taken into account and is shown by Fig. 1. The extrema of local regions in the image at different scales for

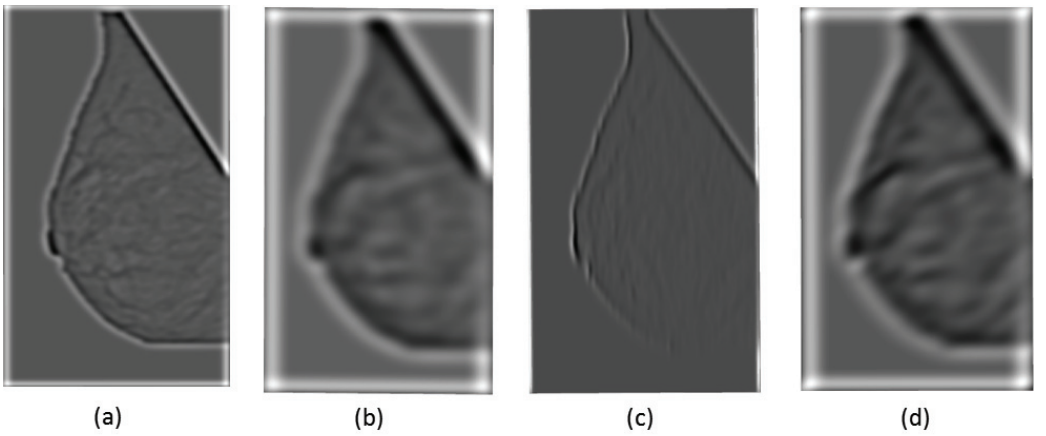

Fig. 3. Example gLoG processing based on the mammogram shown in Fig. 1(b) and the gLoG kernels shown in Fig. 2.

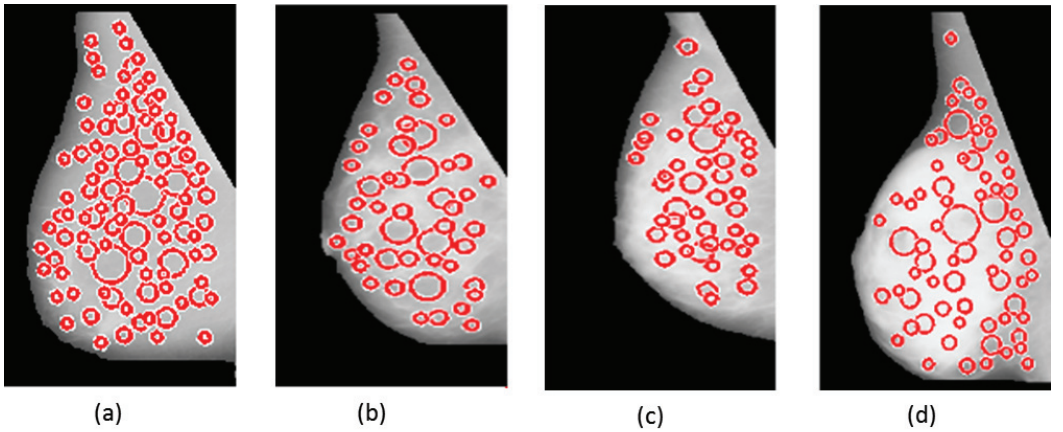

Fig. 4. Blobs detecting fatty tissue at multi-scale after merging based on qualitative relations for mammograms sorted from BIRADS I to BIRADS IV

$\{\sigma_x = 10, \sigma_y = 10\}$ are estimated. Such detected blobs represent the approximate size of blob like fatty and dense tissue regions. To calculate the overall amount of blob like fatty/dense tissue structures from the image, the blobs are identified at multiple scales. All these blobs are merged together to represent the total amount of dense/fatty tissue in the image. The overlapped blobs due to the closely related tissues are removed while merging the blobs at different scales. The overlapped and intersected blobs are removed using the above qualitative relations. The merging procedure starts from the largest scale to the smallest scale. The external blobs are retained. If the distance between the blobs is less than the radius of the largest blob, the inner blob is eliminated. If they are intersecting and if the distance between them is less than the radius of the largest blob, it leads to its deletion. Blob like structures detected for fatty tissues are represented in Fig. 4 while blob like structures detected for dense tissues are represented in Fig. 5

4. Density classification

In order to find the correlation between the BIRADS classes and the dense/fatty areas detected from blob detection methodology, we classified the mammograms from MIAS database on the basis of BIRADS class. The MIAS database was classified into 4 groups of BIRADS classes by an expert radiologist.To classify the MIAS mammograms based on the BIRADS classes, we used the K-nearest neighbour classifier algorithm[17].The K-nearest Neighbour classifier

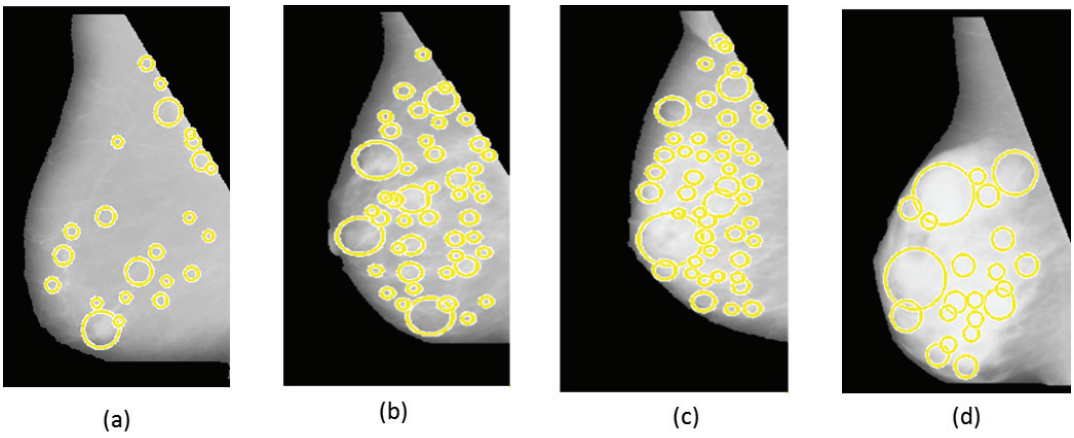

Fig. 5. Blobs detecting dense tissue at multi-scale after merging based on qualitative relations for mammograms sorted from BIRADS I to BIRADS IV

(kNN) classifies a non-classified vector into 'k' most nearest similar vectors from the training set. kNN is based on the distance between the sample points in the feature space data sample. The classification result by the k-NN classifier is represented by Table 1.The MIAS mammographic images are classified into BIRADS classes based on the features obtained from blob detection.

Table 1. Confusion matrix for k-NN classifier based on BIRADS classes

	Automatic Classification			
Truth Data	**Birads I**	**Birads II**	**Birads III**	**Birads IV**
Birads I	44	21	13	9
Birads II	22	37	35	9
Birads III	12	36	35	10
Birads IV	8	9	11	9

To find the correlation between classification by classifier on the BIRADS class and the risk, the MIAS database was divided into low risk and high risk mammograms based on the tissue type they have. BIRADS I and II were considered low risk as they have a low amount of dense tissue while BIRADS III and IV were considered high risk mammograms as they have more dense tissue. The same classifier, k-NN algorithm was used to classify the mammograms as low/high risk. It gave a classification similar to the Birads classification and is illustrated in Table 2.

Table 2. Confusion matrices of k-NN classifier for classifying low/high risk

	Automatic Classification	
	Low Risk	**High risk**
Low Risk	132	60
High risk	65	63

The method was applied on a set of 320 mammograms taken from the MIAS database.To evaluate the results, we used a 10 fold cross validation method on the entire data set. The confusion matrix for the classifier is shown in Table 1. The rows in the confusion matrix indicate the true class and columns indicate the label the classifiers associates

with the object. It was found that the accuracy of k-nn classifier on classifying the mammograms as high-risk and low-risk based on the dense area and fatty area found by the blob detection method is 61%. This is correlated to the results obtained in Table 1 where the mammographic images are classified based on BIRADS version 4[4].

5. Discussion and Conclusions

The main purpose of the paper has been in determining the fatty/dense tissue present in mammograms using multi-scale blob detection method and to classify the mammograms based on the BIRADS classification leading to the classification of mammographic images into high risk and low risk mammograms.

We have shown that the fatty/dense tissue area obtained from the mammographic images through the generalized Laplacian of Gaussian based approach can be used to classify the mammographic images into high/low risk mammograms. The comparison between the classification of BIRADS class and high/low risk mammograms based on the fatty/dense tissue obtained from the multi-scale blob method give a linear relationship.

Future work will focus on classifying the mammograms using more fatty/dense blob structures obtained from a wide number of kernels at different range of ratios of $\{\sigma_x$ and $\sigma_y\}$. In addition, we will evaluate the classification on larger digitised and digital datasets to estimate the relationship between dense/fatty tissue and classify them as low/high risk and find the relation between the Birads classes. Similarly the classification of mammograms into high risk and low risk will be improved through evaluating more features at different orientations.

References

1. Wolfe, John.N,. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, vol.37, pp.2486–2492,(1976).
2. Boyd N.F, Byng J.W, Jong R.A, Fishell E.K, Little L.E, Miller A.B, Lockwood G.A, Tritchler D.L, Yaffe M.J.: Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study.*Journal of the National Cancer Institute*, vol.87,pp. 670–675 (1995).
3. Vachon C.M, Van Gils C.H, Sellers T.A, Ghosh K, Pruthi S, Brandt K.R, Pankratz V.S. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Research*,vol.9(6),pp.1-9, 2007.
4. American College of Radiology. BI-RADS Committee and American College of Radiology: Breast imaging reporting and data system. American College of Radiology (1998).
5. Tabar L ,Tot T, Dean P: The art and science of early detection with mammography. Breast Cancer. Stuttgart, Germany: Thieme (2005).
6. Muhimmah I , Oliver A ,Denton ERE , Pont J , Zwiggelaar R: Comparison between Wolfe, Boyd, BI-RADS and Tabár based mammographic risk assessment. : *Lecture Notes in Computer Science*, vol.4046,pp. 407. Springer, (2006).
7. He W, Denton E , Zwiggelaar R: Mammographic segmentation and risk classification using a novel binary model based bayes classifier. *Breast Imaging*. pp. 40–47 (2012).
8. Chen Z , Zwiggelaar R: A modified fuzzy c-means algorithm for breast tissue density segmentation in mammograms.In: *Information Technology and Applications in Biomedicine (ITAB),2010 10th IEEE International Conference* on (pp. 1-4). IEEE(2010).
9. Chen.Z, Oliver.A, Denton E.R.E., and Zwiggelaar.R: Automated mammographic risk classification based on breast density estimation.In *Pattern Recognition and Image Analysis LNCS*, vol.(7887), pp. 237–244. Springer, (2013).
10. He W, Juette A, Denton E. R. E, Oliver A, Mart R and Zwiggelaar R : A Review on Automatic Mammographic Density and Parenchymal Segmentation. International Journal of Breast Cancer. *International journal of breast cancer*.Vol (2015),(2015)
11. Chen Z, Arnau O, Denton E.R.E, Zwiggelaar R.: A multiscale blob representation of mammographic parenchymal patterns and mammographic risk assessment. In *Computer Analysis of Images and Patterns*Vol. 12(8),pp. 3838-3850 (2013).
12. Illingworth .J, Kittler .J: A survey of the Hough transform.*Computer vision, graphics, and image processing*. vol.44,pp. 87–116. Elsevier (1988).
13. Gonzales R, Woods R, and Eddins.S:Digital image processing using MATLAB. Pearson Education India(2004).
14. Kong H, Akakin H.C, Sarma S.E:A generalized laplacian of gaussian filter for blob detection and its applications. *IEEE Transactions on Cybernetics*. vol.43,pp.1719–1733 (2013).
15. Suckling.J, Parker.J, Dance D. R, Astley S.M, Hutt I, Boggis C.R.M, Ricketts .I, Stamatakis E, Cerneaz N, Kok S.L, Taylor P, Betal D, and Savage J , : The mammographic image analysis society digital mammogram database. In: *Proceedings of International Workshop Digital Mammography*, pp.211–221.(1994).
16. Oliver A., Freixenet J, Mart R, Pont J, Perez E, Denton E, Zwiggelaar R: A novel breast tissue density classification methodology. *IEEE Transactions on Information Technology in Biomedicine*, vol. 12,pp.55-65 (2008)
17. Duda, Richard O, Peter E.H, David G.S. *Pattern classification*. John Wiley and Sons, (2012).
18. George.M, Rampun.A, Denton.E, Zwiggelaar .R :Mammographic Ellipse Modelling Towards Birads Density Classification. In *Proceedings of the 13th International Workshop on Breast Imaging* June 19 - 22, 2016, Malm, Sweden.